January 2016. Horticultural Plant Journal, 2 (1): 16–25.

Horticultural Plant Journal

Available online at www.sciencedirect.com

The journal's homepage: http://www.journals.elsevier.com/horticultural-plant-journal

Evolution of *TWIN SISTER of FT* (*TSF*) Genes in Brassicaceae

HU Yunyan, LIU Bo, SUN Chao, LIU Jing, WANG Xiaobo, CHENG Feng, LIANG Jianli, WANG Xiaowu, and WU Jian *

Institute of Vegetables and Flowers, Chinese Academy of Agricultural Sciences, Beijing 100081, China

Received 6 September 2015; Received in revised form 16 November 2015; Accepted 23 December 2015

Available online 19 February 2016

Abstract

FT and its homolog, *TWIN SISTER OF FT* (*TSF*), act redundantly as integrators in floral transition pathways in *Arabidopsis thaliana*. The evolution of these key flowering regulatory genes during Brassicaceae speciation has not been well studied; therefore, we investigated their evolution in 13 sequenced Brassicaceae species. While the phylogenetic analysis indicated that *FT* gene evolution has followed two independent lineage-specific routes, *TSF* evolution does not appear to have been completely consistent within the Brassicaceae lineage I and lineage II division. The two *TSF* copies in the *Thellungiella* genus were divided into A and B groups in the phylogenetic analysis. Examination of conserved non-coding sequences and conserved domains within a 5 kb region upstream of the *TSF* start codon revealed the same group division inferred by the phylogenetic analysis. In addition, *TSF* genes retained syntenic relationships among genes in the same group, but not between group A and group B. The two copies of the *TSF* gene in the *Thellungiella* species were syntenic to the *TSF* genes in group A and group B, respectively. We also identified *TSF-A* gene residues in the syntenic region of group B species, but no *TSF-B* residues could be found in the group A syntenic region. Therefore, we inferred that the *TSF* genes in lineage II species experienced a duplication event after diversification from lineage I. Following their split from *Thellungiella*, *Brassica* species lost the ancestral *TSF* gene and retained the duplicated copy.

Keywords: Brassicaceae; *FT*; *TSF*; evolution; CNS; synteny

1. Introduction

In flowering plants, the switch from vegetative to reproductive growth is a major developmental transition that is highly responsive to various environmental cues (Koornneef et al., 2004). Controlling the timing of this transition is especially critical to ensure high agricultural productivity in crop plants. More than 180 genes that control flowering time have been identified by analysis of transgenic plants or by isolation of loss-of-function mutants in the model plant species *Arabidopsis thaliana* (Fornara et al., 2010). In *A. thaliana*, six major pathways that control flowering time have been identified: the photoperiod, vernalization, autonomous, gibberellin, ambient temperature and age pathways (Roux et al., 2006; Alonso-Blanco et al., 2009; Kim et al., 2009; Fornara et al., 2010).

The photoperiod pathway controls flowering time by promoting flowering during long summer days and repressing it during short winter days (Fornara et al., 2010). The *FT* gene acts as a floral promoter, mainly in the photoperiod pathway downstream of *CONSTANS* (*CO*). *FT* is expressed in the distal portion of the leaf and encodes a protein that is transported via the phloem to the meristem (Kardailsky et al., 1999; Turck et al., 2008). The floral repressor *FLOWERING LOCUS C* (*FLC*) represses *FT* transcription and integrates the autonomous and vernalization pathways (Samach et al., 2000; Hepworth et al., 2002). CO promotes flowering by initiating the transcription of *FT* and *TWIN SISTER OF FT* (*TSF*) genes (Fornara et al., 2010). *TSF* is the closest homolog of *FT* in *A. thaliana*, with 82% identity between their amino acid sequences (Yamaguchi et al., 2005). Overexpression of *TSF* causes a precocious flowering phenotype similar to a phenotype resulting from overexpression of *FT* (Kobayashi et al., 1999). Thus, *FT* and *TSF* may have similar roles in the promotion of the floral transition. These two genes have a similar tissue preference — expression in the phloem —

* Corresponding author. Tel.: +86 10 82105971.

E-mail address: wujian@caas.cn

Peer review under responsibility of Chinese Society for Horticultural Science (CSHS) and Institute of Vegetables and Flowers (IVF), Chinese Academy of Agricultural Sciences (CAAS)

http://dx.doi.org/10.1016/j.hpj.2016.02.001

but distinct spatial patterns: *TSF* is mainly expressed in hypocotyls, whereas *FT* is primarily expressed in cotyledons and leaves (Yamaguchi et al., 2005).

The Brassicaceae family, which contains many important horticultural and agricultural crops, has been classified into two major groups: the early branching *Aethionema* group and the core Brassicaceae group (Franzke et al., 2011; Haudry et al., 2013). The core group comprises three major lineages (I, II, and III) (Beilstein et al., 2006). To date, the complete genomes of five lineage I species have been sequenced: *A. thaliana* (Initiative, 2000), *A. lyrata* (Hu et al., 2011), *Capsella rubella* (Slotte et al., 2013), *Leavenworthia alabamica* (Haudry et al., 2013) and *Camelina sativa* (Kagale et al., 2014). Seven species from lineage II have been sequenced: *Brassica rapa* (Wang et al., 2011), *Thellungiella parvula* (Dassanayake et al., 2011), *T. salsuginea* (Wu et al., 2012), *T. halophila* (Yang et al., 2013), *Sisymbrium irio* (Haudry et al., 2013), *B. oleracea* (Liu et al., 2014) and *B. napus* (Chalhoub et al., 2014). No species in lineage III have been sequenced thus far. In recent years, the evolution of Brassicaceae species has been studied. Haudry et al. (2013) constructed a phylogenetic tree of nine Brassicaceae species with 1 048 889 four-fold-degenerate sites and *Carica papaya* as an outgroup. Zhang et al. (2015a, 2015b) clarified the lineage-specific evolution of Methylthioalkylmalate (MAM) synthase genes involved in glucosinolates biosynthesis of 13 sequenced Brassicaceae species. Brassicaceae genomes comprise 24 basic genomic blocks (GBs), A–X (Schranz et al., 2006). The common ancestor of *B. rapa* and *B. oleracea* underwent a whole-genome triplication process; therefore, three copies of each GB have been identified (Wang et al., 2011). The three GB copies differ with regard to the extent of their gene loss (fractionation), with *Brassica* GBs classified into least fractionated (LF), medium fractionated (MF1), and most fractionated (MF2) subgenomes, according to their gene density (Wang et al., 2011). The *B. napus* (genome $A_nA_nC_nC_n$) is the product of hybridization between *B. rapa* (genome A_rA_r) and *B. oleracea* (genome C_oC_o), followed by chromosome doubling; this process is known as allopolyploidization (Chalhoub et al., 2014).

Plant conserved non-coding sequences (CNSs) are a specific category of phylogenetic footprint (Freeling and Subramaniam, 2009). Over evolutionary time, non-functional sequences are expected to diverge faster than sequences under selective constraints. CNSs are considered to be functional elements because they are not removed by purifying selection and genetic drift (Freeling and Subramaniam, 2009). Plant CNSs are enriched in transcription factors or *cis*-acting binding sites, and are usually clustered around genes. Their relatively small genome sizes, robust phylogeny, and wealth of genomic data make Brassicaceae an ideal family for the identification of CNSs (Haudry et al., 2013). An atlas of over 90 000 CNSs has been established for Brassicaceae (Haudry et al., 2013), providing a basis for analyzing CNS distribution in specific genes.

Chromosomal synteny analysis is important to reveal the genomic evolution of related species (Tang et al., 2008), such as those in Brassicaceae. Syntenic genes from different species are orthologs located on syntenic fragments; therefore, they often share similar functions and originate from a common ancestor (Lyons et al., 2008; Cheng et al., 2012a).

In this study, we investigated the evolution of *FT* and *TSF* genes in 13 sequenced Brassicaceae species by constructing a phylogenetic tree of *FT* and *TSF* genes, analyzing the distribution of CNSs and conserved domains within a 5 kb region upstream of the *TSF* start codon, and examining the syntenic relationship of the *TSF* genes.

2. Materials and methods

2.1. Sources of genome data

Genomic datasets for *A. thaliana*, *A. lyrata*, and *B. rapa* were downloaded from TAIR (https://www.arabidopsis.org/index.jsp), the Joint Genome Initiative database (http://genome.jgi-psf.org/Araly1/Araly1.home.html), and BRAD (http://brassicadb.org/brad/; Cheng et al., 2011), respectively. Gene and genome data for 10 Brassicaceae species were obtained from published studies as follows: *A. arabicum*, *L. alabamica* and *S. irio* (Haudry et al., 2013); *C. rubella* (Slotte et al., 2013); *C. sativa* (Kagale et al., 2014); *T. parvula* (Dassanayake et al., 2011); *T. salsuginea* (Wu et al., 2012); *T. halophila* (Yang et al., 2013); *B. oleracea* (Liu et al., 2014); and *B. napus* (Chalhoub et al., 2014).

2.2. FT and TSF gene identification

Orthologous *FT* and *TSF* genes in Brassicaceae species were identified by analyzing their syntenic relationships with *A. thaliana* using a genome-wide BLASTP search. We first searched for syntenic orthologs of *FT* (AT1G65480) and *TSF* (AT4G20370) in other Brassicaceae species on BRAD using two criteria: protein sequence similarity (*E*-value ≤1.0E-20) and collinearity of flanking genes (Cheng et al., 2012a). Taking into account the possibility that not all *FT* and *TSF* genes in Brassicaceae species share a syntenic relationship with their orthologs in *A. thaliana*, we used protein sequences of *A. thaliana FT* and *TSF* to identify their orthologs in other species by BLASTP searching, with an *E*-value ≤1.0E-20 and sequence coverage ≥70% (Altschul et al., 1997). The *E*-value and sequence coverage of *FT* and *TSF* genes in 13 sequenced Brassicaceae species are listed in Table 1.

2.3. Phylogenetic analysis of TSF genes

The MUSCLE program was used, with default parameters, to align the full-length sequences of FT and TSF proteins in the 13 sequenced Brassicaceae species (Edgar, 2004). A phylogenetic tree of *FT* and *TSF* genes was constructed using the maximum likelihood method, as implemented in MEGA 5.0, with 1 000 bootstrap replicates (Tamura et al., 2011).

2.4. Identification of CNSs and conserved domains

We analyzed CNSs and conserved domains present in the 5 kb region upstream from the start codon of each *TSF* gene. CNSs located in the upstream region of *AtTSF* and *AlTSF* were extracted from the CNS dataset reported by Haudry et al. (2013). The extracted CNSs of *AlTSF* were aligned to the corresponding region of other Brassicaceae *TSF* genes using BLASTN.

Table 1 The *E*-value and sequence coverage of *FT* and *TSF* genes in 13 sequenced Brassicaceae species

Species	Gene ID	*E*-value	Sequence coverage /%
A. arabicum	AA_scaffold6005_7	6.0E-90	89
C. rubella	Carubv10021034m	3.0E-96	94
	Carubv10007441m	1.0E-82	81
L. alabamica	LA_scaffold230_6	3.0E-93	90
	LA_scaffold1919_11	2.0E-94	90
	LA_scaffold3390_1	4.0E-95	90
	LA_scaffold1783_27	6.0E-79	79
C. sativa	Csa07g028300	3.0E-96	94
	Csa16g022760	1.0E-95	94
	Csa05g068740	1.0E-96	95
	Csa11g025850	1.0E-81	81
	Csa12g038920	4.0E-66	91
A. lyrata	Al_scaffold_201177.1	2.0E-99	96
	Al_scaffold_0007_2115	2.0E-87	87
T. parvula	c0021_00055	1.0E-86	85
	c0006_00976	2.0E-80	79
	c0013_00348	1.0E-77	78
T. halophila	Thhalv10019214m	1.0E-86	87
	Thhalv10027204m	5.0E-79	79
	Thhalv10019213m	2.0E-78	77
T. salsuginea	Tsa5g21060	1.0E-86	87
	Tsa7g19290	5.0E-79	79
	Tsa5g34510	2.0E-78	77
S. irio	SI_scaffold658_12	9.0E-88	85
	SI_scaffold108_38	1.0E-80	82
B. rapa	Bra022475	1.0E-87	86
	Bra004117	4.0E-85	82
	Bra015710	5.0E-74	83
B. oleracea	Bol012573	3.0E-85	81
	Bol045330	2.0E-47	86
	Bol027595	2.0E-39	70
	Bol017639	1.0E-63	75
B. napus	GSBRNA2T00124448001	8.0E-85	82
	GSBRNA2T00090951001	3.0E-87	86
	GSBRNA2T00067517001	1.0E-39	73
	GSBRNA2T00146560001	4.0E-74	76
	GSBRNA2T00077948001	2.0E-64	75
	GSBRNA2T00113342001	6.0E-75	77

Table 2 *FT* and *TSF* genes identified in 13 sequenced Brassicaceae species

Species	Gene ID	Reference
A. arabicum	*AaFT* (AA_scaffold6005_7)	Haudry et al., 2013
C. rubella	*CrFT* (Carubv10021034m)	Slotte et al., 2013
	CrTSF (Carubv10007441m)	
L. alabamica	*LaFT-1* (LA_scaffold230_6)	Haudry et al., 2013
	LaFT-2 (LA_scaffold1919_11)*	
	LaFT-3 (LA_scaffold3390_1)*	
	LaTSF (LA_scaffold1783_27)	
C. sativa	*CsFT-1* (Csa07g028300)	Kagale et al., 2014
	CsFT-2 (Csa16g022760)	
	CsFT-3 (Csa05g068740)	
	CsTSF-1 (Csa11g025850)	
	CsTSF-2 (Csa12g038920)	
A. thaliana	*AtFT* (AT1G65480)	Initiative, 2000
	AtTSF (AT4G20370)	
A. lyrata	*AlFT* (Al_scaffold_201177.1)	Hu et al., 2011
	AlTSF (Al_scaffold_0007_2115)	
T. parvula	*TpFT* (c0021_00055)	Dassanayake et al., 2011
	TpTSF-1 (c0006_00976)	
	TpTSF-2 (c0013_00348)	
T. halophila	*ThFT* (Thhalv10019214m)	Yang et al., 2013
	ThTSF-1 (Thhalv10027204m)	
	ThTSF- 2(Thhalv10019213m)	
T. salsuginea	*TsFT* (Tsa5g21060)	Wu et al., 2012
	TsTSF-1 (Tsa7g19290)	
	TsTSF-2 (Tsa5g34510)	
S. irio	*SiFT* (SI_scaffold658_12)	Haudry et al., 2013
	SiTSF (SI_scaffold108_38)	
B. rapa	*BrFT-1* (Bra022475)	Wang et al., 2011
	BrFT-2 (Bra004117)	
	BrTSF (Bra015710)	
B. oleracea	*BoFT-1* (Bol012573)	Liu et al., 2014
	BoFT-2 (Bol045330)	
	BoTSF-1 (Bol027595)	
	BoTSF-2 (Bol017639)*	
B. napus	*BnFT-1* (GSBRNA2T00124448001)	Chalhoub et al., 2014
	BnFT-2 (GSBRNA2T00090951001)	
	BnFT-3 (GSBRNA2T00067517001)	
	BnTSF-1 (GSBRNA2T00146560001)	
	BnTSF-2 (GSBRNA2T00077948001)*	
	BnTSF-3 (GSBRNA2T00113342001)*	

Note: Asterisk indicates non-syntenic genes.

Aligned hits satisfying syntenic relationship criteria were manually selected and treated as candidate CNSs.

To identify conserved domains, any repeats occurring more than 50 times in the entire genome sequence were masked in each species (Turco et al., 2013). We extracted the upstream sequences of *TSF* genes from the masked genome and used *TpTSF-1* and *TpTSF-2* as references to align with *TSF* genes from groups A and B, respectively. 'Bl2Seq' was used to align the upstream sequences and filter the aligned hits, using the parameters of Turco et al. (2013). Finally, we manually selected the aligned hits satisfying the syntenic relationship criteria and considered them to be candidate conserved domains.

2.5. Analysis of TSF gene residues

The full coding sequences (CDSs) of *TpTSF-1* and *TpTSF-2* were used to identify residues of the ancestral *TSF* gene in group A and B species (excluding *T. parvula*, *T. salsuginea*, and *T. halophila*) by genome-wide BLASTN searching, respectively. The residues were identified according to sequence length (≥15 bp), sequence similarity (identity ≥ 80%), and sequence location (located within 10 kb of *TSF* non-syntenic intergenic regions). Residues present in intergenic regions were re-annotated using SoftBerry (http://linux1.softberry.com/).

3. Results

3.1. Identification of FT and TSF genes in Brassicaceae genomes

FT orthologs in the other analyzed Brassicaceae species were identified by their syntenic relationships and sequence similarities. Eight species possessed one *FT* copy (*A. arabicum*, *A. thaliana*, *C. rubella*, *A. lyrata*, *T. parvula*, *T. salsuginea*, *T. halophila* and *S. irio*). Two species, *B. rapa* and *B. oleracea*, retained two *FT* copies, and three species, *L. alabamica*, *C. sativa*, and *B. napus*, contained three. Detailed information on the *FT* genes in the 13 sequenced Brassicaceae species is presented in Table 2.

In *A. thaliana*, *TWIN SISTER OF FT* (*TSF*) is the closest homolog of *FT* and acts as a floral pathway integrator redundantly with *FT* (Kobayashi et al., 1999; Yamaguchi et al., 2005). We applied the same method used to search for *FT* orthologs to identify *TSF* orthologs in the 13 Brassicaceae species. We found that six species, *A. thaliana*, *C. rubella*, *L. alabamica*, *A. lyrata*, *S. irio*, and *B. rapa*, possessed one *TSF* copy, while five species, *T. parvula*, *T. salsuginea*, *T. halophila*, *B. oleracea*, and *C. sativa*, retained two. Only one species, *B. napus*, contained three *TSF* copies. Table 2 lists the details of the *TSF* genes in the 13 sequenced Brassicaceae species.

3.2. Independent evolution of FT and TSF genes in core Brassicaceae

Taking the *Carica papaya FT* gene as an outgroup, we constructed a phylogenetic tree of *FT* and *TSF* genes, based on the protein sequences of the 13 sequenced species (Fig. 1). The phylogenetic tree contained two major clades representing the *FT* and *TSF* genes, with the *FT* gene of *A. arabicum* falling outside of these two clades. The *FT* clade (Fig. 1, blue branch) was divided into groups C and D, corresponding to lineage I and lineage II, respectively. Group C included all three *FT* copies from *L. alabamica* and *C. sativa* and the single copies from *A. lyrata*, *A. thaliana*, and *C. rubella*. Group D comprised both duplicated *FT* genes from each of the three *Brassica* species, as well as the single copy from the three *Thellungiella* species and *S. irio*.

In the case of *TSF* genes, two groups (A and B) were also present. One copy of the *TSF* gene of *T. parvula*, *T. salsuginea*, and *T. halophila* was found in each group. In addition to a copy of the *TSF-1* gene from each of the three *Thellungiella* species (*TpTSF-1*, *ThTSF-1*, and *TsTSF-1*), group A included *AtTSF*, two *TSF* genes from *C. sativa*, and one copy of *TSF* gene of *L. alabamica*, *A. lyrata*, *C. rubella*, and *S. irio*. *TSF* genes from

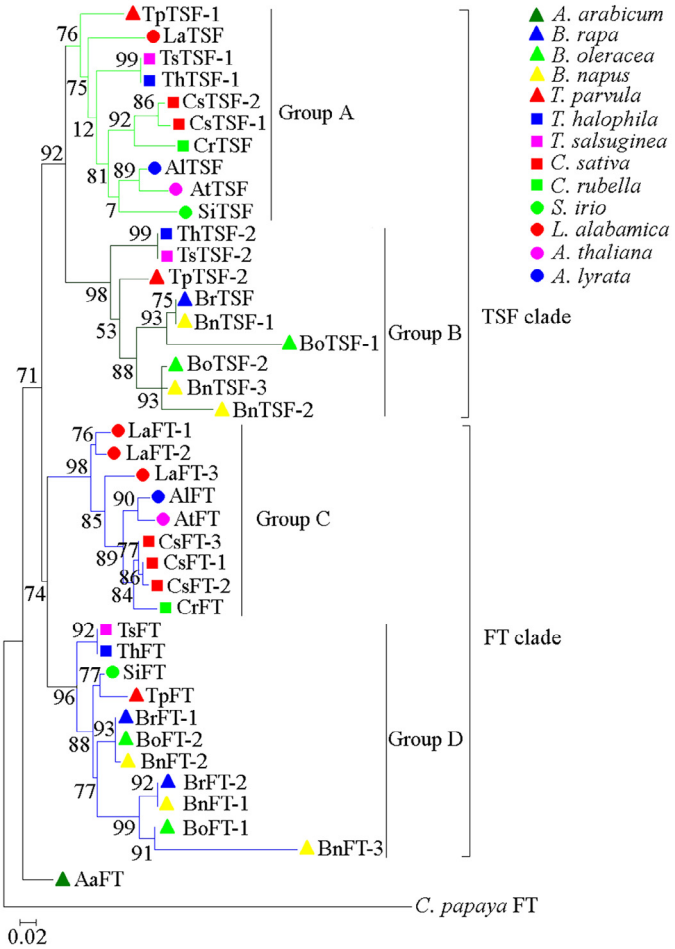

Fig. 1 Phylogenetic relationships of *FT* and *TSF* genes in Brassicaceae, based on their protein sequences

The phylogenetic tree was constructed using the full-length sequences of FT and TSF proteins from 13 sequenced Brassicaceae species: *Aethionema arabicum*, *Arabidopsis thaliana*, *A. lyrata*, *Capsella rubella*, *Leavenworthia alabamica*, *Camelina sativa*, *Thellungiella parvula*, *T. salsuginea*, *T. halophila*, *Sisymbrium irio*, *Brassica rapa*, *B. oleracea* and *B. napus*. The FT protein of *Carica papaya* was used as an outgroup. Numbers on branches are bootstrap support percentages based on 1 000 replicates. The scale bar is 0.02 substitutions per site. Genes in the core Brassicaceae group formed two major clades (*FT* and *TSF*). The *TSF* and *FT* clades were divided into two groups (*TSF*: A and B; *FT*: C and D)

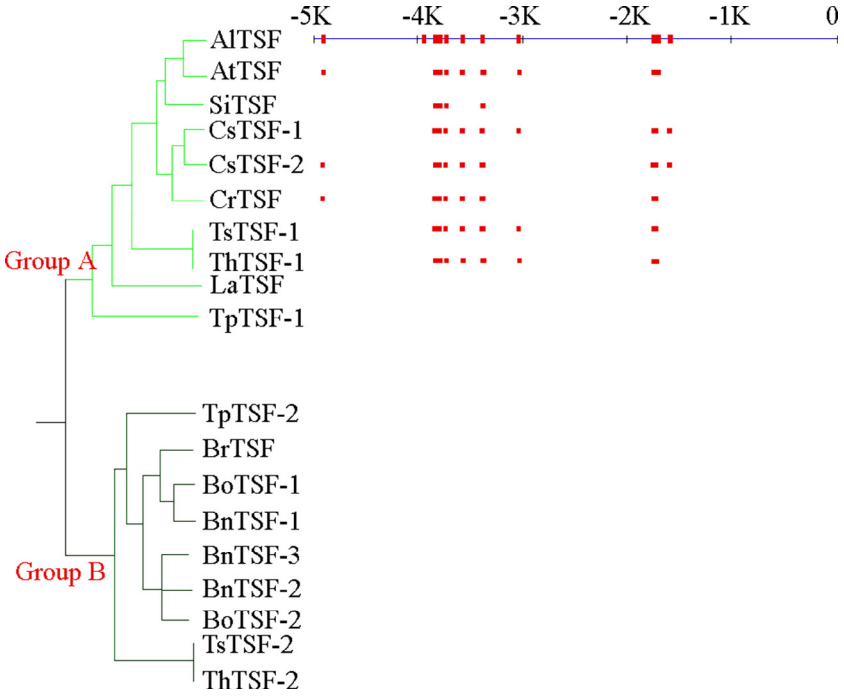

Fig. 2 Distribution of conserved non-coding sequences (CNS) in groups A and B

The phylogenetic tree of *TSF* genes is shown on the left and the CNS distribution is depicted on the right. The red rectangles are CNS regions. Numbers above the rectangles are kb positions of CNSs upstream of the *TSF* start codon. *AlTSF* was used as a reference for both groups A and B

B. rapa (*BrTSF*), *B. oleracea* (*BoTSF-1* and *BoTSF-2*) and *B. napus* (*BnTSF-1*, *BnTSF-2*, *BnTSF-3*) were clustered together with each of the *TSF-2* genes from three *Thellungiella* species (*TpTSF-2*, *ThTSF-2*, *TsTSF-2*) in group B.

On the basis of the inferred phylogenic relationships, we formulated two hypotheses. The first hypothesis is that two *TSF* genes were present in the ancestral Brassicaceae genome. Over the course of evolution, the *Thellungiella* genus kept both genes, whereas lineage I species and the *Brassica* genus each retained one of the two different copies. Our second hypothesis is that a single *TSF* gene existed in the ancestral Brassicaceae genome. The lineage I species inherited this copy. In lineage II, however, the *TSF* gene underwent a duplication event after the diversification from lineage I. After their split from *Thellungiella*, the *Brassica* species lost the ancestral *TSF* copy and retained the duplicated one.

3.3. Different distributions of CNSs and conserved domains between groups A and B

We investigated the distributions of CNSs and conserved domains situated within the 5 kb region upstream of the start codon of *TSF* genes. No CNSs were present in this region in *AtTSF*; therefore, we used the CNS distribution of *AlTSF* as a reference to analyze *TSFs* from the other species. CNS distributions in the 5 kb upstream region were very different between group A and group B genes. Most *TSF* genes in group A shared a similar CNS distribution, although *LaTSF* and *TpTSF-1* had no CNSs in the upstream region (Fig. 2). In contrast to group A, no CNSs were present in this region in group B *TSFs*.

TpTSF-1 and *TpTSF-2* were used as references to investigate the diversity of conserved domains within and between groups A and B (Fig. 3). The distributions of conserved domains located 0 to 400 bp upstream of the start codon were similar between the two groups. The distributions of more distant conserved domains, however, barely overlapped between group A and group B. In group A, conserved domains were distributed within the 1 to 1.5 kb upstream region; in group B, they were located 1.3 to 3 kb upstream. Thus, the distribution of conserved domains has diversified between groups A and B.

3.4. TSF gene expression levels in B. rapa, B. oleracea, and Arabidopsis thaliana

We investigated the expression levels of *TSF* genes from groups A and B by searching available transcriptome data from *B. rapa* and *B. oleracea*. No *BrTSF* gene expression was detected using *B. rapa* transcriptome data generated by Cheng et al. (2012b) from three organs (root, stem, and leaf). Similarly, a search of *B. oleracea* gene expression data from seven different tissues (bud, callus, root, stem, leaf, flower, and silique) in the GEO database (accession number GSE42891) revealed that *BoTSF* was not expressed. Yamaguchi et al. (2005) have reported that *TSF* in *A. thaliana* is mainly expressed in hypocotyls of seedlings, and in the flowers and developing siliques of mature plants. These expression data suggested that there might be expressional divergence of *TSF* genes between groups A and B.

3.5. The syntenic relationship of TSF genes between groups A and B and evolutionary patterns of TSF genes among Brassicaceae species

The common *Brassica* ancestor has been confirmed to have undergone a whole-genome triplication (Wang et al., 2011),

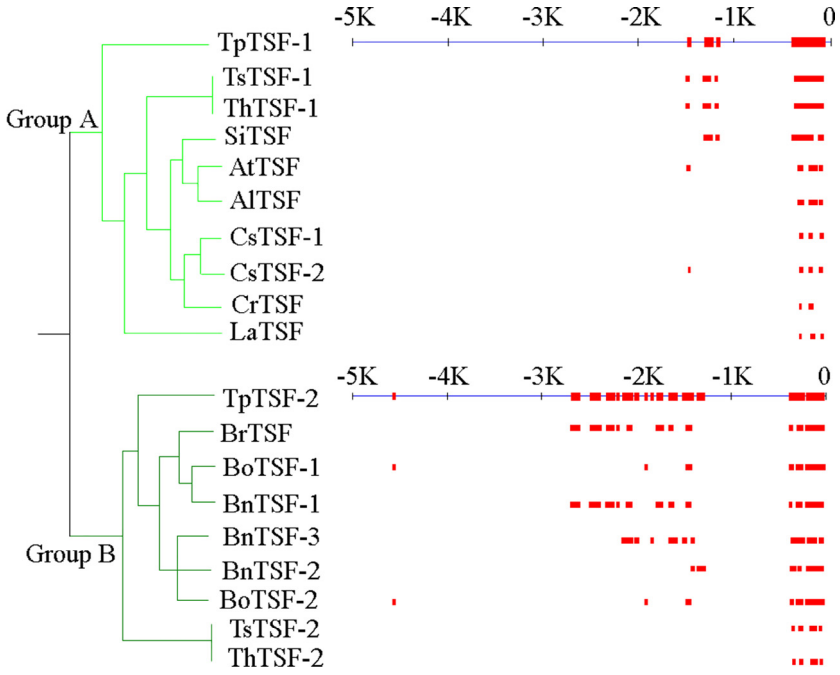

Fig. 3 Distribution of conserved domains in groups A and B. The phylogenetic tree of *TSF* genes is shown on the left and the conserved domain distribution is displayed on the right

The red rectangles are conserved domains. Numbers above the rectangles are kb positions of conserved domains upstream of the *TSF* start codon. *TpTSF-1* and *TpTSF-2* were used as references for groups A and B, respectively

thereby giving rise to three subgenomes (LF, MF1, and MF2) (Wang et al., 2011). In the present study, all *TSF* genes in the genus *Brassica* were found on the LF subgenome. Of the 24 conserved GBs in Brassicaceae species (Schranz et al., 2006), *TSF* genes from group A (*TSF-A*) were found on block U, while those from group B (*TSF-B*) were located on block E (Fig. 4). Within each group, *TSF* genes shared good syntenic relationships. In contrast, no syntenic relationship was observed between *TSF-A* and *TSF-B*.

To determine whether either of our two hypotheses (I or II) could account for *TSF* gene evolution in Brassicaceae, we aligned the CDS sequence of *TpTSF-1* with the region where *TSF-A* syntenic genes were missing in each group B species, including the three *Brassica* species. As a result, residues of *TpTSF-1* were found in this region in all three *Brassica* species (Fig. 4). In the *B. rapa* genome, the intergenic region lacking the *TSF-A* syntenic genes was highly conserved with respect to the *TpTSF-1* gene, with 21% sequence coverage (112 of 528 bp) and 93% identity in the covered region (Table 3). In the *B. oleracea* genome, *TpTSF-1* sequence coverage in the intergenic region was 14% (75 of 528 bp) with 96% identity (Table 3). *B. napus* is an allotetraploid plant with an A genome from *B. rapa* and a C genome from *B. oleracea*. In the LF subgenome of A and C genomes, *TSF-A* gene residues were found with *TpTSF-1* sequence coverages of 21% (112 of 528 bp; 93% identity) and 14% (75 of 528 bp; 96% identity), respectively (Table 3). Taking into account the possibility of incomplete gene annotations of the *Brassica* genomes, we re-annotated the intergenic region marked by missing *TSF-A* syntenic genes in *Brassica*. We found no additional genes in this region in *B. rapa*, *B. oleracea*, or *B. napus*. Using the same alignment method detailed above for group B species, we looked for *TSF-B* gene residues in regions where syntenic genes of *TpTSF-2* were missing in *A. thaliana*, *A. lyrata*, *S. irio*, *L. alabamica*, *C. rubella* and *C. sativa* (Fig. 4). In contrast to our findings with respect to group B and *TSF-A* genes, no *TSF-B* residues were detected in these group A species.

On the basis of the above results, we propose the evolutionary model for *TSF* genes in Brassicaceae shown in Fig. 5. According to this model, only one copy of the *TSF* gene was present in the Brassicaceae ancestral genome. The ancestral *TSF* gene was duplicated in lineage II species after their divergence from lineage I. Lineage I species (*L. alabamica*, *A. thaliana*, *A. lyrata*, *C. rubella, C. sativa*, and *S. irio*) possessed the original copy. Following the split of *Brassica* and *Thellungiella* in lineage II, *Brassica* species, including *B. rapa*, *B. oleracea*, and *B. napus*, retained the duplicated *TSF* and lost the ancestral one (Fig. 5).

Table 3 Information on *TSF-A* residues in group B species

Species	Subgenome	Total sequence length/bp	Total sequence coverage	Average sequence identity
B. rapa	LF	112	21%	93%
B. oleracea	LF	75	14%	96%
B. napus	A-LF	112	21%	93%
	C-LF	75	14%	96%

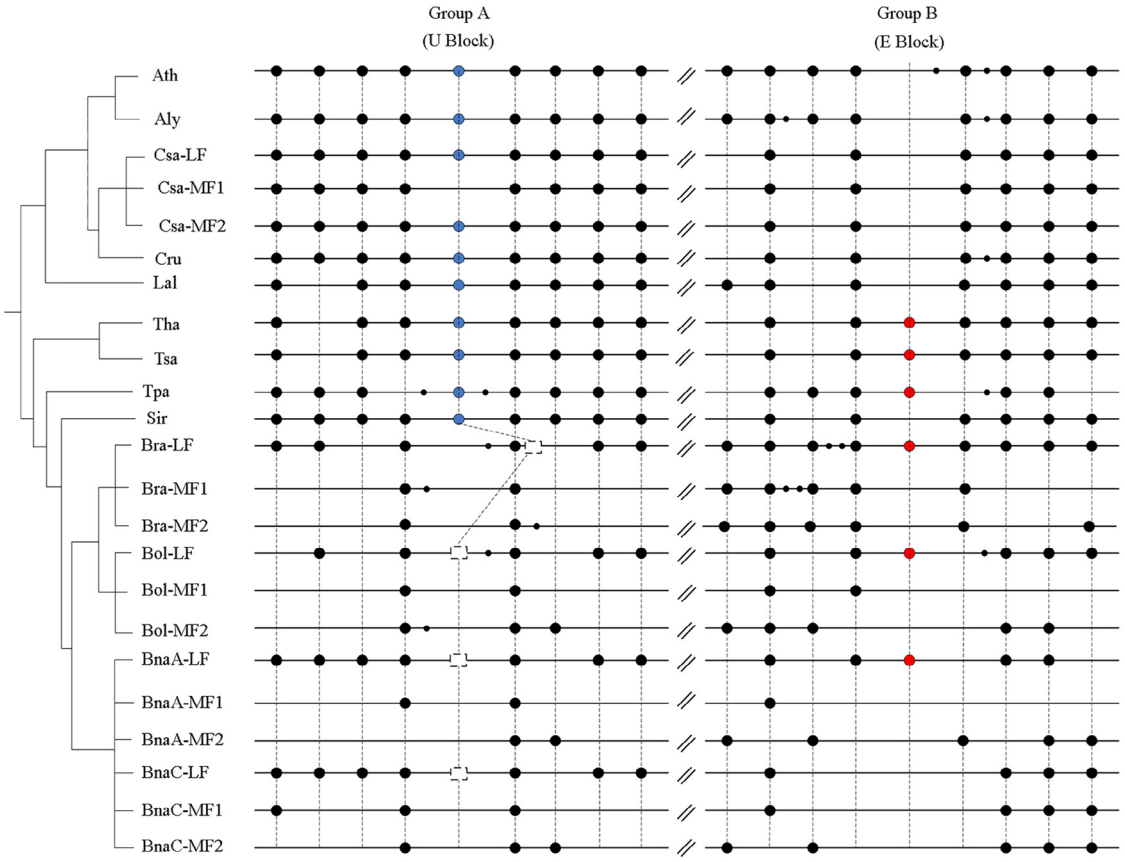

Fig. 4 Synteny analysis of *TSF* regions in Brassicaceae showing group-specific rearrangements. The phylogenetic tree on the left shows the evolutionary relationships of 13 sequenced Brassicaceae species

The *TSF* region exhibits conserved synteny across the group A sequenced Brassicaceae genomes, as well as across group B genomes. *TSF* genes in groups A and B are represented by blue and red dots, respectively. The small black dots indicate non-syntenic genes, and the dashed rectangles refer to residues of *TpTSF-1* in the genus *Brassica*

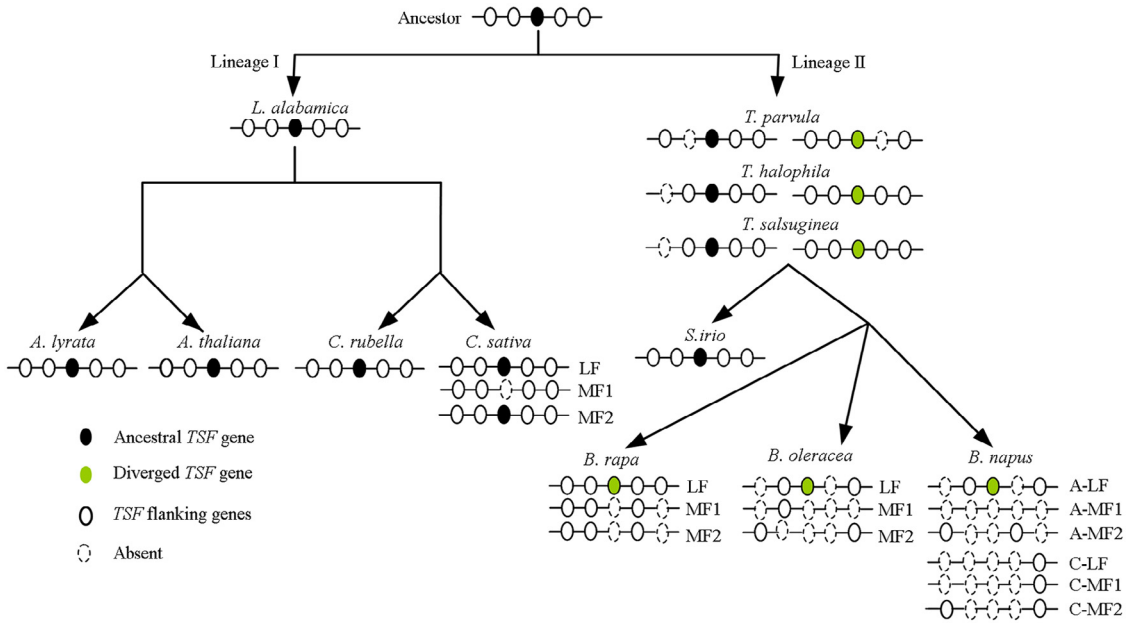

Fig. 5 Proposed model of *TSF* gene evolution in Brassicaceae

In this model, *TSF* gene evolution proceeded via two independent routes that were specific to lineages I and II. The black and green circles represent *TSF* genes; the white circles are flanking genes

4. Discussion

Multiple flowering time control pathways are integrated by *FT* to control flowering (Moon et al., 2005), with *FT* being the direct target of CO, FLC, and SVP (SHORT VEGETATIVE PHASE) proteins (Samach et al., 2000; Searle et al., 2006; Li et al., 2008). In *A. thaliana*, *TSF* is the homologous gene of *FT* and shares a similar role in promoting the floral transition (Yamaguchi et al., 2005). A genome-wide association study of flowering time variation in *A. thaliana* showed that the important role of *TSF* has been underestimated under greenhouse conditions (Brachi et al., 2010). No report on *TSF* function in other Brassicaceae species has appeared thus far, and whether this gene has the same flowering regulatory function as in *A. thaliana* remains unknown. In preparation for a future analysis of *FT* and *TSF* functions in Brassicaceae species, in this study we took advantage of the 13 available Brassicaceae genome sequences to investigate the evolution and diversification of *FT* and *TSF* genes in crucifers.

Brassica is closely related to *A. thaliana*. The genus underwent a genome triplication event that preceded the origin of the diploid species *Brassica rapa* (AA), *B. oleracea* (CC), and *B. nigra* (BB) (Wang et al., 2011). Mesopolyploidy has had contrasting effects on genes involved in flowering time control: multiple copies of *FLC* and *VERNALIZATIO N1* (*VRN1*) produced by whole-genome triplication have been preserved in the *B. rapa* genome, whereas *GI*, *SHORT VEGETATIVE PHASE* (*SVP*), and the three *CONSTANS-LIKE* (*COL*) genes are limited to only one copy (Wang et al., 2011). Our work on *B. rapa* showed that in general, the genes involved in flowering time regulation were over-retained (Wu et al., 2012), while no syntenic ortholog of *AtTSF* was found in the *B. rapa* genome.

Our phylogenetic analysis of *FT* and *TSF* genes from the 13 sequenced Brassicaceae species divided both *FT* and *TSF* genes into two clades. *A. arabicum* is part of an early branching sister group to the core Brassicaceae (Haudry et al., 2013). The *TSF* gene exists in the core Brassicaceae group, but not in *A. arabicum*. This absence suggested that *FT* and *TSF* may not have diverged before the split of *A. arabicum* and the core Brassicaceae group. The phylogenetic tree of *FT* genes in the 13 sequenced Brassicaceae species is consistent with previously published phylogenies based on either genomic blocks (Bailey et al., 2006; Schranz et al., 2007) or CNS analysis (Haudry et al., 2013). This similarity indicates that *FT* genes have evolved in a lineage-specific pattern in Brassicaceae. *TSF* evolution has differed from that of *FT* genes, however, and is not completely consistent with the lineage division. Although *TSF* genes were separated into two groups, similar to *FT* genes, the two *TSF* copies of the three *Thellungiella* species were divided into different groups. Interestingly, we observed a diversification of the CNS distribution between *TSF* genes of *T. halophila*, *T. salsuginea*, and *T. parvula.* The CNS distribution in the two *TSF* copies of *T. halophila* and *T. salsuginea* followed the same pattern as the other members of the group to which they belonged. In contrast, the CNS distribution in both *TSF* copies of *T. parvula* (*TpTSF-1* and *TpTSF-2*) resembled that of group B. *T. parvula* has been identified previously as the diploid ancestor of *Brassica* species (Cheng et al., 2013). The distribution of CNSs within the 5 kb region upstream of *TpTSF-1* and *TpTSF-2* is in line with this identification. The uncovered syntenic relationships between groups A and B, and the presence of *TSF* residues in the intergenic region characterized by missing *TSF* syntenic genes, helped verify our second hypothetical model of *TSF* gene evolution in the Brassicaceae family.

The diversification of intergenic regions raised the question of whether *TSF* genes in groups A and B would have different expression patterns. Our search of the available transcriptome data of *B. rapa* and *B. oleracea* identified no detectable expression of *TSF* in *B. rapa* or *B. oleracea*, whereas *AtTSF* is mainly expressed in seedling hypocotyls, mature plant flowers and developing siliques (Yamaguchi et al., 2005). This difference suggested that *TSFs* are functionally diversified between groups A and B. The availability of additional transcriptome data from other Brassicaceae species would help to confirm this speculation.

Among the 13 species, the three *Brassica* species are the most widely cultivated crops. *B. napus* (genome $A_nA_nC_nC_n$) was formed by a hybridization between *B. rapa* (genome A_rA_r) and *B. oleracea* (genome C_oC_o) (Chalhoub et al., 2014). Thus, we expected to find two syntenic *TSFs* in the *B. napus* genome; however, we only identified one syntenic *TSF* copy (*BnTSF-1*), present on the LF subgenome derived from the A_r genome. No copy was found on the C_o genome. When we carefully examined the region from which the *BnTSF* gene was absent, we found a large 60 kb gap. Thus, the absence of an identified syntenic copy on the C_o genome could reflect an incomplete genome assembly of *B. napus*. *S. irio*, a species from lineage II, also underwent a duplication event after the diversification from lineage I. *S. irio* retained the ancestral *TSF* gene after its split from *Thellungiella* and lost the duplicated copy. We expected to observe residues of the duplicated copy in the intergenic region of the missing *TSF-B* syntenic gene, but found none. A detailed examination of the region in which the duplicated *SiTSF* was missing uncovered a large gap. Consequently, the absence of *SiTSF* duplicated copy residues may also be due to an incomplete *S. irio* genome assembly.

Acknowledgements

We are grateful to Jifang Zhang, Zhiyuan Liu and Baozhen Gao for their helpful suggestions. The work was funded by the National High Technology R&D Program of China (2012AA100101), the National Program on Key Basic Research Projects of China (The 973 Program, 2012CB113906), and by NSFC funding (31272179). The research was carried out in the Key Laboratory of Biology and Genetic Improvement of Horticultural Crops, Ministry of Agriculture, P. R. China, and at the Sino-Dutch Joint Lab of Horticultural Genomics Technology in Beijing.

References

Alonso-Blanco, C., Aarts, M.G.M., Bentsink, L., Keurentjes, J.J.B., Matthieu, R., Vreugdenhil, D., Koornneef, M., 2009. What has natural variation taught

us about plant development, physiology, and adaptation? Plant Cell, 21: 1877–1896.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res, 25: 3389–3402.

Bailey, C.D., Koch, M.A., Mayer, M., Mummenhoff, K., O'Kane, S.L., Jr., Warwick, S.I., Windham, M.D., Al-Shehbaz, I.A., 2006. Toward a global phylogeny of the Brassicaceae. Mol Biol Evol, 23: 2142–2160.

Beilstein, M.A., Al-Shehbaz, I.A., Kellogg, E.A., 2006. Brassicaceae phylogeny and trichome evolution. Am J Bot, 93: 607–619.

Brachi, B., Faure, N., Horton, M., Flahauw, E., Vazquez, A., Nordborg, M., Bergelson, J., Cuguen, J., Roux, F., 2010. Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. PLoS Genet, 6: e1000940.

Chalhoub, B., Denoeud, F., Liu, S., Parkin, I.A., Tang, H., Wang, X., Chiquet, J., Belcram, H., Tong, C., Samans, B., Corréa, M., Da Silva, C., Just, J., Falentin, C., Koh, C.S., Le Clainche, I., Bernard, M., Bento, P., Noel, B., Labadie, K., Alberti, A., Charles, M., Arnaud, D., Guo, H., Daviaud, C., Alamery, S., Jabbari, K., Zhao, M., Edger, P.P., Chelaifa, H., Tack, D., Lassalle, G., Mestiri, I., Schnel, N., Le Paslier, M.C., Fan, G., Renault, V., Bayer, P.E., Golicz, A.A., Manoli, S., Lee, T.H., Thi, V.H., Chalabi, S., Hu, Q., Fan, C., Tollenaere, R., Lu, Y., Battail, C., Shen, J., Sidebottom, C.H., Wang, X., Canaguier, A., Chauveau, A., Bérard, A., Deniot, G., Guan, M., Liu, Z., Sun, F., Lim, Y.P., Lyons, E., Town, C.D., Bancroft, I., Wang, X., Meng, J., Ma, J., Pires, J.C., King, G.J., Brunel, D., Delourme, R., Renard, M., Aury, J.M., Adams, K.L., Batley, J., Snowdon, R.J., Tost, J., Edwards, D., Zhou, Y., Hua, W., Sharpe, A.G., Paterson, A.H., Guan, C., Wincker, P., 2014. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. Science, 345: 950–953.

Cheng, F., Liu, S., Wu, J., Fang, L., Sun, S., Liu, B., Li, P., Hua, W., Wang, X., 2011. BRAD, the genetics and genomics database for *Brassica* plants. BMC Plant Biol, 11: 136.

Cheng, F., Mandáková, T., Wu, J., Xie, Q., Lysak, M.A., Wang, X., 2013. Deciphering the diploid ancestral genome of the mesohexaploid *Brassica rapa*. Plant Cell, 25: 1541–1554.

Cheng, F., Wu, J., Fang, L., Wang, X., 2012a. Syntenic gene analysis between *Brassica rapa* and other Brassicaceae species. Front Plant Sci, 3: 198.

Cheng, F., Wu, J., Fang, L., Sun, S., Liu, B., Lin, K., Bonnema, G., Wang, X., 2012b. Biased gene fractionation and dominant gene expression among the subgenomes of *Brassica rapa*. PLoS ONE, 7: e36442.

Dassanayake, M., Oh, D.H., Haas, J.S., Hernandez, A., Hong, H., Ali, S., Yun, D.J., Bressan, R.A., Zhu, J.-K., Bohnert, H.J., Cheeseman, J.M., 2011. The genome of the extremophile crucifer *Thellungiella parvula*. Nat Genet, 43: 913–918.

Edgar, R.C., 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res, 32: 1792–1797.

Fornara, F., Montaigu, A., Coupland, G., 2010. SnapShot: Control of flowering in *Arabidopsis*. Cell, 550: e551–e552.

Franzke, A., Lysak, M.A., Al-Shehbaz, I.A., Koch, M.A., Mummenhoff, K., 2011. Cabbage family affairs: The evolutionary history of Brassicaceae. Trends Plant Sci, 16: 108–116.

Freeling, M., Subramaniam, S., 2009. Conserved noncoding sequences (CNSs) in higher plants. Curr Opin Plant Biol, 12: 126–132.

Haudry, A., Platts, A.E., Vello, E., Hoen, D.R., Leclercq, M., Williamson, R.J., Forczek, E., Joly-Lopez, Z., Steffen, J.G., Hazzouri, K.M., Dewar, K., Stinchcombe, J.R., Schoen, D.J., Wang, X., Schmutz, J., Town, C.D., Edger, P.P., Pires, J.C., Schumaker, K.S., Jarvis, D.E., Mandáková, T., Lysak, M.A., van den Bergh, E., Schranz, M.E., Harrison, P.M., Moses, A.M., Bureau, T.E., Wright, S.I., Blanchette, M., 2013. An atlas of over 90,000 conserved noncoding sequences provides insight into crucifer regulatory regions. Nat Genet, 45: 891–898.

Hepworth, S.R., Valverde, F., Ravenscroft, D., Mouradov, A., Coupland, G., 2002. Antagonistic regulation of flowering-time gene *SOC1* by *CONSTANS* and *FLC* via separate promoter motifs. EMBO J, 21: 4327–4337.

Hu, T.T., Pattyn, P., Bakker, E.G., Cao, J., Cheng, J.F., Clark, R.M., Fahlgren, N., Fawcett, J.A., Grimwood, J., Gundlach, H., Haberer, G., Hollister, J.D., Ossowski, S., Ottilar, R.P., Salamov, A.A., Schneeberger, K., Spannagl, M., Wang, X., Yang, L., Nasrallah, M.E., Bergelson, J., Carrington, J.C., Gaut, B.S., Schmutz, J., Mayer, K.F.X., van de Peer, Y., Grigoriev, I.V., Nordborg, M., Weigel, D., Guo, Y.L., 2011. The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. Nat Genet, 43: 476–481.

Initiative, A.G., 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature, 408: 796–815.

Kagale, S., Koh, C., Nixon, J., Bollina, V., Clarke, W.E., Tuteja, R., Spillane, C., Robinson, S.J., Links, M.G., Clarke, C., Higgins, E.E., Huebert, T., Sharpe, A.G., Parkin, I.A.P., 2014. The emerging biofuel crop *Camelina sativa* retains a highly undifferentiated hexaploid genome structure. Nat Commun, 5: 3706.

Kardailsky, I., Shukla, V.K., Ahn, J.H., Dagenais, N., Christensen, S.K., Nguyen, J.T., Chory, J., Harrison, M.J., Weigel, D., 1999. Activation tagging of the floral inducer *FT*. Science, 286: 1962–1965.

Kim, D.H., Doyle, M.R., Sung, S., Amasino, R.M., 2009. Vernalization: Winter and the timing of flowering in plants. Annu Rev Cell Dev Biol, 25: 277–299.

Kobayashi, Y., Kaya, H., Goto, K., Iwabuchi, M., Araki, T., 1999. A pair of related genes with antagonistic roles in mediating flowering signals. Science, 286: 1960–1962.

Koornneef, M., Alonso-Blanco, C., Vreugdenhil, D., 2004. Naturally occurring genetic variation in *Arabidopsis thaliana*. Annu Rev Plant Biol, 55: 141–172.

Li, D., Liu, C., Shen, L., Wu, Y., Chen, H., Robertson, M., Helliwell, C.A., Ito, T., Meyerowitz, E., Yu, H., 2008. A repressor complex governs the integration of flowering signals in *Arabidopsis*. Dev Cell, 15: 110–120.

Liu, S., Liu, Y., Yang, X., Tong, C., Edwards, D., Parkin, I.A., Zhao, M., Ma, J., Yu, J., Huang, S., Wang, X., Wang, J., Lu, K., Fang, Z., Bancroft, I., Yang, T.J., Hu, Q., Wang, X., Yue, Z., Li, H., Yang, L., Wu, J., Zhou, Q., Wang, W., King, G.J., Pires, J.C., Lu, C., Wu, Z., Sampath, P., Wang, Z., Guo, H., Pan, S., Yang, L., Min, J., Zhang, D., Jin, D., Li, W., Belcram, H., Tu, J., Guan, M., Qi, C., Du, D., Li, J., Jiang, L., Batley, J., Sharpe, A.G., Park, B.S., Ruperao, P., Cheng, F., Waminal, N.E., Huang, Y., Dong, C., Wang, L., Li, J., Hu, Z., Zhuang, M., Huang, Y., Huang, J., Shi, J., Mei, D., Liu, J., Lee, T.H., Wang, J., Jin, H., Li, Z., Li, X., Zhang, J., Xiao, L., Zhou, Y., Liu, Z., Liu, X., Qin, R., Tang, X., Liu, W., Wang, Y., Zhang, Y., Lee, J., Kim, H.H., Denoeud, F., Xu, X., Liang, X., Hua, W., Wang, X., Wang, J., Chalhoub, B., Paterson, A.H., 2014. The *Brassica oleracea* genome reveals the asymmetrical evolution of polyploid genomes. Nat Commun, 5: 3930.

Lyons, E., Pedersen, B., Kane, J., Alam, M., Ming, R., Tang, H., Wang, X., Bowers, J., Paterson, A., Lisch, D., Freeling, M., 2008. Finding and comparing syntenic regions among *Arabidopsis* and the outgroups papaya, poplar, and grape: CoGe with rosids. Plant Physiol, 148: 1772–1781.

Moon, J., Lee, H., Kim, M., Lee, I., 2005. Analysis of flowering pathway integrators in *Arabidopsis*. Plant Cell Physiol, 46: 292–299.

Roux, F., Touzet, P., Cuguen, J., Corre, V.L., 2006. How to be early flowering: An evolutionary perspective. Trends Plant Sci, 11: 375–381.

Samach, A., Onouchi, H., Gold, S.E., Ditta, G.S., Schwarz-Sommer, Z., Yanofsky, M.F., Coupland, G., 2000. Distinct roles of CONSTANS target genes in reproductive development of *Arabidopsis*. Science, 288: 1613–1616.

Schranz, M.E., Lysak, M.A., Mitchell-Olds, T., 2006. The ABC's of comparative genomics in the Brassicaceae: Building blocks of crucifer genomes. Trends Plant Sci, 11: 535–542.

Schranz, M.E., Song, B.H., Windsor, A.J., Mitchell-Olds, T., 2007. Comparative genomics in the Brassicaceae: A family-wide perspective. Curr Opin Plant Biol, 10: 168–175.

Searle, I., He, Y., Turck, F., Vincent, C., Fornara, F., Krober, S., Amasino, R.A., Coupland, G., 2006. The transcription factor *FLC* confers a flowering response to vernalization by repressing meristem competence and systemic signaling in *Arabidopsis*. Genes Dev, 20: 898–912.

Slotte, T., Hazzouri, K.M., Agren, J.A., Koenig, D., Maumus, F., Guo, Y.L., Steige, K., Platts, A.E., Escobar, J.S., Newman, L.K., Wang, W., Mandáková, T., Vello, E., Smith, L.M., Henz, S.R., Steffen, J., Takuno, S., Brandvain, Y., Coop, G., Andolfatto, P., Hu, T.T., Blanchette, M., Clark, R.M., Quesneville, H., Nordborg, M., Gaut, B.S., Lysak, M.A., Jenkins, J., Grimwood, J., Chapman, J., Prochnik, S., Shu, S., Rokhsar, D., Schmutz, J., Weigel, D., Wright, S.I., 2013. The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. Nat Genet, 45: 831–835.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol, 28: 2731–2739.

Tang, H., Bowers, J.E., Wang, X., Ming, R., Alam, M., Paterson, A.H., 2008. Synteny and collinearity in plant genomes. Science, 320: 486–488.

Turck, F., Fornara, F., Coupland, G., 2008. Regulation and identity of florigen: *FLOWERING LOCUS T* moves center stage. Annu Rev Plant Biol, 59: 573–594.

Turco, G., Schnable, J.C., Pedersen, B., Freeling, M., 2013. Automated conserved non-coding sequence (CNS) discovery reveals differences in gene content and promoter evolution among grasses. Front Plant Sci, 4: 170.

Wang, X.W., Wang, H.Z., Wang, J., Sun, R., Wu, J., Liu, S., Bai, Y., Mun, J.H., Bancroft, I., Cheng, F., Huang, S., Li, X., Hua, W., Wang, J., Wang, X., Freeling, M., Pires, J.C., Paterson, A.H., Chalhoub, B., Wang, B., Hayward, A., Sharpe, A.G., Park, B.S., Weisshaar, B., Liu, B., Li, B., Liu, B., Tong, C., Song, C., Duran, C., Peng, C., Geng, C., Koh, C., Lin, C., Edwards, D., Mu, D., Shen, D., Soumpourou, E., Li, F., Fraser, F., Conant, G., Lassalle, G., King, G.J., Bonnema, G., Tang, H., Wang, H., Belcram, H., Zhou, H., Hirakawa, H., Abe, H., Guo, H., Wang, H., Jin, H., Parkin, I.A., Batley, J., Kim, J.S., Just, J., Li, J., Xu, J., Deng, J., Kim, J.A., Li, J., Yu, J., Meng, J., Wang, J., Min, J., Poulain, J., Wang, J., Hatakeyama, K., Wu, K., Wang, L., Fang, L., Trick, M., Links, M.G., Zhao, M., Jin, M., Ramchiary, N., Drou, N., Berkman, P.J., Cai, Q., Huang, Q., Li, R., Tabata, S., Cheng, S., Zhang, S., Zhang, S., Huang, S., Sato, S., Sun, S., Kwon, S.J., Choi, S.R., Lee, T.H., Fan, W., Zhao, X., Tan, X., Xu, X., Wang, Y., Qiu, Y., Yin, Y., Li, Y., Du, Y., Liao, Y., Lim, Y., Narusaka, Y., Wang, Y., Wang, Z., Li, Z., Wang, Z., Xiong, Z., Zhang, Z., Brassica rapa Genome Sequencing Project Consortium, 2011. The genome of the mesopolyploid crop species *Brassica rapa*. Nat Genet, 43: 1035–1157.

Wu, H.J., Zhang, Z., Wang, J.Y., Oh, D.H., Dassanayake, M., Liu, B., Huang, Q., Sun, H.X., Xia, R., Wu, Y., Wang, Y.N., Yang, Z., Liu, Y., Zhang, W., Zhang, H., Chu, J., Yan, C., Fang, S., Zhang, J., Wang, Y., Zhang, F., Wang, G., Lee, S.Y., Cheeseman, J.M., Yang, B., Li, B., Min, J., Yang, L., Wang, J., Chu, C., Chen, S.Y., Bohnert, H.J., Zhu, J.K., Wang, X.J., Xie, Q., 2012. Insights into salt tolerance from the genome of *Thellungiella salsuginea*. Proc Natl Acad Sci U S A, 109: 12219–12224.

Yamaguchi, A., Kobayashi, Y., Goto, K., Abe, M., Araki, T., 2005. *TWIN SISTER OF FT* (*TSF*) acts as a floral pathway integrator redundantly with *FT*. Plant Cell Physiol, 46: 1175–1189.

Yang, R.L., Jarvis, D.E., Chen, H., Beilstein, M.A., Grimwood, J., Jenkins, J., Shu, S.Q., Prochnik, S., Xin, M.M., Ma, C., Schmutz, J., Wing, R.A., Mitchell-Olds, T., Schumaker, K.S., Wang, X.F., 2013. The reference genome of the halophytic plant *Eutrema salsugineum*. Front Plant Sci, 4: 46.

Zhang, J.F., Wang, X.B., Cheng, F., Wu, J., Liang, J.L., Yang, W.C., Wang, X.W., 2015a. Lineage-specific evolution of Methylthioalkylmalate synthases (MAMs) involved in glucosinolates biosynthesis. Front Plant Sci, 6: 18.

Zhang, X., Meng, L., Liu, B., Hu, Y., Cheng, F., Liang, J., Aarts, M.A.M., Wang, X., Wu, J., 2015b. A transposon insertion in *FLOWERING LOCUS T* is associated with delayed flowering in *Brassica rapa*. Plant Sci, 241: 211–220.